CHAPTER 3

REGRESSION AND THE LINEAR MODEL

The aim of **regression modelling** is to explain the observed variation in a **response** variable, Y, by relating it to some collection of **predictor** variables, $X_1, X_2, ..., X_D$. The variation is decomposed into a **systematic** component, that is, a deterministic function of the predictor variables, and a **random** component that is the result of measurement procedures (or unmeasurable variability). The simplest model is the **Normal Linear Model**, where the systematic component is a function of the predictors and some model parameters, β, and the random variation is assumed to be the result of additive normally distributed random error terms. This model is explained in section 3.1.

An extension of this framework is necessary if non-normal data are to modelled. In section **??**, a generalization of the linear model framework is introduced, and this leads us to the **Generalized Linear Model (GLM)** framework. We shall see that the statistical ideas and procedures of likelihood, estimation and testing introduced previously are applicable to the GLM framework.

3.1   THE NORMAL LINEAR MODEL

We assume that the variables to be modelled are as follows; we will observe paired data, with response data y_i paired to predictor variables stored in vector form $x_i = (x_{i1}, ..., x_{iD})^T$, and our aim is to explain the variation in $(y_1, ..., y_n)$. We achieve this by modelling the conditional distribution of response variable Y_i given the observed value of predictor variable $X_i = x_i$. Specifically, we may write

$$Y_i = \beta_0 + \beta_1 x_1 + ... + \beta_D x_{iD} + \varepsilon_i = \beta_0 + \sum_{j=1}^{D} \beta_j x_{ij} + \varepsilon_i \tag{3.1.1}$$

where $\varepsilon_i \sim N\left(0, \sigma^2\right)$ for $i = 1, ...n$ are independent and identically distributed random error terms. Note that this implies

$$Y_i | X_i = x_i \sim N\left(\beta_0 + \sum_{j=1}^{D} \beta_j x_{ij}, \sigma^2\right) \qquad \therefore \qquad E_{f_{Y|X}}\left[Y_i | X_i = x_i\right] = \beta_0 + \sum_{j=1}^{D} \beta_j x_{ij}. \tag{3.1.2}$$

In vector notation, (**??**) can be re-written $Y_i = x_i^T \beta + \varepsilon_i$, where $x_i = (1, x_{i1}, x_{i2}, ..., x_{iD})^T$, and thus, for vector $Y = (Y_1, ..., Y_n)^T$ we have

$$Y = \mathbf{X}\beta + \varepsilon$$

where $\mathbf{X}$ is a $n \times (D + 1)$ matrix called the **design** matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1D} \\ 1 & x_{21} & \cdots & x_{2D} \\ 1 & x_{31} & \cdots & x_{3D} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nD} \end{bmatrix}$$

and to mimic (3.1.2)

$$Y \sim N_n \left(\mathbf{X}\beta, \sigma^2 I_n \right) \tag{3.1.3}$$

where I_n is the $n \times n$ identity matrix, giving a joint pdf for Y given $\mathbf{X}$ of the form

$$f_{Y|\beta,\sigma^2}(y; \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mathbf{X}\beta)^T (y - \mathbf{X}\beta) \right\} \tag{3.1.4}$$

3.1.1   THE EXTENDED LINEAR MODEL

The formulation of the linear model above can be extended to allow for more general dependence on the predictors.   Suppose that $g_1, g_2, ..., g_K$ are K (potentially non-linear) functions of the D original predictors, that is

$$g_k(x_i) = g_k(x_{i1}, ..., x_{iD})$$

is some scalar function, for example, we could have

- $g_k(x_{i1}, ..., x_{iD}) = g_k(x_{i1}) = x_{i1}$ (the identity function)

- $g_k(x_{i1}, ..., x_{iD}) = g_k(x_{i1}) = a_k \sqrt{x_{i1}}$

- $g_k(x_{i1}, ..., x_{iD}) = g_k(x_{i1}) = a_k \log x_{i1}$

- $g_k(x_{i1}, ..., x_{iD}) = g_k(x_{i1}, x_{i2}) = a_k x_{i1} + b_k x_{i2}$

and so on.   This reformulation does not effect our probabilistic definition of the model in (3.1.3); we can simply redefine design matrix $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} 1 & g_1(x_1) & \cdots & g_K(x_1) \\ 1 & g_1(x_2) & \cdots & g_K(x_2) \\ 1 & g_1(x_3) & \cdots & g_K(x_3) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & g_1(x_n) & \cdots & g_K(x_n) \end{bmatrix}$$

now an $n \times (K+1)$ matrix.   In the discussion below, we will regard the **transformed** variables $(g_1(X), g_2(X), ..., g_K(X))$ as the predictors and drop the dependence on the transformation functions. Hence we have

- Y as a $n \times 1$ column vector

- $\mathbf{X}$ as a $n \times (K+1)$ matrix with ith row $(1, g_1(x_i), ..., g_K(x_i))$

- β as a $(K+1) \times 1$ column vector

3.1.2   MAXIMUM LIKELIHOOD ESTIMATION IN THE LINEAR MODEL

Maximum likelihood estimation for the normal linear model is straightforward. Recall that if $\theta = \left(\beta, \sigma^2 \right)$ then the mle $\hat{\theta}$ is given by

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} f_{Y|\beta,\sigma^2}(y; \beta, \sigma^2) = \arg \max_{\theta \in \Theta} L\left(\beta, \sigma^2; y, x \right)$$

where parameter space $\Theta \equiv \mathbb{R}^K \times \mathbb{R}^+$. Taking logs in (3.1.4) gives

$$\log L\left(\beta, \sigma^2; y, x\right) = -\frac{n}{2}\log \sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\left(y - \mathbf{X}\beta\right)^T\left(y - \mathbf{X}\beta\right) \qquad (3.1.5)$$

and considering the maximization for β indicates

$$\arg\max_{\beta \in \mathbb{R}^K} \log L\left(\beta, \sigma^2; y, x\right) = \arg\min_{\beta \in \mathbb{R}^K}\left(y - \mathbf{X}\beta\right)^T\left(y - \mathbf{X}\beta\right)$$

and thus,

$$S\left(\beta\right) = \left(y - \mathbf{X}\beta\right)^T\left(y - \mathbf{X}\beta\right) = y^T y - y^T\mathbf{X}\beta - \beta^T\mathbf{X}^T y + \beta^T\mathbf{X}^T\mathbf{X}\beta = y^T y - 2y^T\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta.$$

Using vector/matrix differentiation

$$\frac{d}{d\beta}\left\{y^T\mathbf{X}\beta\right\} = y^T\mathbf{X} \qquad\qquad \frac{d}{d\beta}\left\{\beta^T\mathbf{X}^T\mathbf{X}\beta\right\} = 2\mathbf{X}^T\mathbf{X}\beta \qquad (3.1.6)$$

and so if $\widehat{\beta}$ is the solution of

$$\frac{dS\left(\beta\right)}{d\beta} = -y^T\mathbf{X} + \mathbf{X}^T\mathbf{X}\beta = 0$$

then it follows that $\widehat{\beta}$ satisfies

$$\mathbf{X}^T\mathbf{X}\widehat{\beta} = \mathbf{X}^T y. \qquad (3.1.7)$$

If the matrix $\mathbf{X}^T\mathbf{X}$ is non-singular, then we have the ML estimates of β as

$$\widehat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y \qquad (3.1.8)$$

and substituting back into (3.1.5) gives

$$\hat{\sigma}^2 = \frac{1}{n}\left(y - \mathbf{X}\widehat{\beta}\right)^T\left(y - \mathbf{X}\widehat{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \qquad (3.1.9)$$

where $\hat{y}_i = x_i^T\widehat{\beta}$ is the **fitted value,** and $y_i - \hat{y}_i$ is the residual. Note that $\mathbf{X}^T\mathbf{X}$ is a symmetric matrix. The expression $\left(y - \mathbf{X}\widehat{\beta}\right)^T\left(y - \mathbf{X}\widehat{\beta}\right)$ is termed the **residual sum of squares** (or **RSS**). A common **adjusted** estimate is

$$\hat{\sigma}^2_{ADJ} = \frac{1}{n - K - 1}\left(y - \mathbf{X}\widehat{\beta}\right)^T\left(y - \mathbf{X}\widehat{\beta}\right) \qquad (3.1.10)$$

the justification for this result depends on the sampling distribution of the estimator. It can be shown that $\hat{\sigma}^2_{ADJ}$ is unbiased for σ^2.

If $K = 1$, with identity function $g(t) = t$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \qquad\qquad \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{1}{\hat{\sigma}^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

and so

$$\widehat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y = \frac{1}{\hat{\sigma}^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} = \frac{1}{\hat{\sigma}^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \\ n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \end{bmatrix}$$

3.1.3   PROPERTIES OF THE ML ESTIMATORS

By elementary properties of random variables, the properties of ML estimator $T = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y$

$$
\begin{aligned}
E_{Y|X,\beta,\sigma^2}[T] &= E_{Y|X,\beta,\sigma^2}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y\right] = \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)E_{Y|X,\beta,\sigma^2}[Y] \\
&= \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\mathbf{X}\beta = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\right)\beta = \beta
\end{aligned}
$$

so that T is unbiased for β, and

$$
\begin{aligned}
Var_{Y|X,\beta,\sigma^2}[T] &= Var_{Y|X,\beta,\sigma^2}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y\right] \\
&= \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)Var_{Y|X,\beta,\sigma^2}[Y]\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)^T \\
&= \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\sigma^2 I_n\left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right) \\
&= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}.
\end{aligned}
$$

Note that, in fact, given β and σ^2

$$
Y \sim N_n\left(\mathbf{X}\beta, \sigma^2 I_n\right) \Longrightarrow T = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y \sim N_{K+1}\left(\beta, \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right). \tag{3.1.11}
$$

It also follows that

$$
(y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta) = \left(y - \mathbf{X}\widehat{\beta}\right)^T\left(y - \mathbf{X}\widehat{\beta}\right) + \left(\widehat{\beta} - \beta\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\widehat{\beta} - \beta\right)
$$

or

$$
S(\beta) = S\left(\widehat{\beta}\right) + \left(\widehat{\beta} - \beta\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\widehat{\beta} - \beta\right)
$$

where

$$
S(\beta) = (y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta) \tag{1}
$$

$$
S\left(\widehat{\beta}\right) = \left(y - \mathbf{X}\widehat{\beta}\right)^T\left(y - \mathbf{X}\widehat{\beta}\right) = (y - \hat{y})^T(y - \hat{y}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}
$$

$$
\left(\widehat{\beta} - \beta\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\widehat{\beta} - \beta\right) = \left(\mathbf{X}\beta - \mathbf{X}\widehat{\beta}\right)^T\left(\mathbf{X}\beta - \mathbf{X}\widehat{\beta}\right) \tag{3}
$$

are the **(1) TOTAL** , **(2) RESIDUAL** and **(3) FITTED** sum of squares (**TSS, RSS** and **FSS).** Therefore, by normal distribution theory, it follows that

$$
\frac{S(\beta)}{\sigma^2} \sim \chi_n^2 \qquad\qquad \frac{S\left(\widehat{\beta}\right)}{\sigma^2} \sim \chi_{n-K-1}^2
$$

so that

$$
s^2 = \frac{S\left(\widehat{\beta}\right)}{(n - K - 1)} \text{ is an } \textbf{UNBIASED} \text{ estimator of } \sigma^2
$$

and the quantity

$$\frac{\widehat{\beta} - \beta}{s.e.\left(\widehat{\beta}\right)} = \frac{\widehat{\beta} - \beta}{s\sqrt{v_{ii}}} \sim Student\left(n - K - 1\right).$$

It also follows that

$$\frac{S\left(\beta\right) - S\left(\widehat{\beta}\right)}{\sigma^2} = \frac{\left(\widehat{\beta} - \beta\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\widehat{\beta} - \beta\right)}{\sigma^2} \sim \chi^2_{K+1}$$

so that finally

$$\frac{\left[S\left(\beta\right) - S\left(\widehat{\beta}\right)\right]/(K+1)}{S\left(\widehat{\beta}\right)/(n - K - 1)} \sim Fisher\left(K + 1, n - K - 1\right)$$

It follows that in this case the ML estimator is the Minimum Variance Unbiased Estimator (MVUE) and the Best Linear Unbiased Estimator (BLUE).

3.1.4   THE ANALYSIS OF VARIANCE

Analysis of variance or **ANOVA** is used to display the sources of variability in a collection of data samples. The ANOVA F-test compares variability **between** samples with the variability **within** samples. In the above analysis, we have that

$$S\left(\beta\right) = S\left(\widehat{\beta}\right) + \left(\widehat{\beta} - \beta\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\widehat{\beta} - \beta\right) \qquad \text{or} \qquad TSS = RSS + FSS.$$

Now, using the distributional results above, we can construct the following **ANOVA Table** to test the hypothesis

$$H_0 : \beta_1 = ... = \beta_K = 0$$

against the general alternative that H_0 is not true.

Source of Variation	D.F.	Sum of squares	Mean square	F
FITTED	K	FSS	FSS/K	$\dfrac{FSS/K}{RSS/(n - K - 1)}$
RESIDUAL	$n - K - 1$	RSS	$RSS/(n - K - 1)$	
TOTAL	$n - 1$	TSS		

This test allows a comparison of the fits of the two competing models implied by the null and alternative hypotheses. Under the null model, if H_0 is true, then the model has $Y_i \sim N\left(\beta_0, \sigma_0^2\right)$ for $i = 1, 2, ...n$, for some β_0 and σ_0^2 to be estimated. Under the alternative hypothesis, there are a total of $K + 1$ β parameters to be estimated using equation (3.1.8). The **degrees of freedom** column headed (D.F.) details how many parameters are used to describe the amount of variation in the corresponding row of the table; for example, for the FIT row, D.F. equals K as there are K parameters used to extend the null model to the alternative model.

Now consider the following design; suppose that there are K possible medical treatments and you wish to test for any difference between them. The parameter vector is $\beta = [\beta_1, \beta_2, ..., \beta_K]^T$ say, and the null hypothesis is that, for some β,

$$H_0 : \beta_1 = \beta_2 = ... = \beta_K = \beta$$

Suppose that there are $n_1, ..., n_K$ observations in the K treatment groups respectively. Then the design matrix in the corresponding (full) linear model takes the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \\ \mathbf{X}_K \end{bmatrix} \qquad \mathbf{X}_k = \left. \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}}_{K \text{ columns}} \right\} n_k \text{ rows}$$

that is, $\mathbf{X}_k$ is a $n_k \times K$ block matrix with only the kth column non-zero, and equal to the $n_k \times 1$ vector of 1s. Under the assumption that the observed responses are normally distributed **with common variance** σ^2 we are in the linear model framework, and all of the above likelihood and statistical theory applies.

3.1.5   ONE-WAY ANOVA

The two sample T-test can be extended to allow a test for differences between more than two data samples. Suppose there are K samples of sizes $n_1, ..., n_K$ from different populations. The model can be represented as follows: let y_{kj} be the jth observation in the kth sample, then

$$y_{kj} = \beta_k + \varepsilon_{kj}$$

for $k = 1, ..., K$, and $\varepsilon_{kj} \sim N\left(0, \sigma^2\right)$. This model assumes that

$$Y_{kj} \sim N\left(\beta_k, \sigma^2\right)$$

and that the expectations for the different samples are different. We can view the data as a table comprising K columns, with each column corresponding to a sample.

To test the hypothesis that each population has the same mean, that is, the hypotheses

$$\begin{aligned} H_0 &: \quad \beta_1 = \beta_2 = ... = \beta_K \\ H_1 &: \quad \text{not } H_0 \end{aligned}$$

an **Analysis of Variance (ANOVA)** F-test may be carried out.

To carry out a test of the hypothesis, the following **ANOVA table** should be completed;

Source	D.F.	Sum of squares	Mean square	F
BETWEEN TREATMENTS	$K-1$	FSS	$FSS/(K-1)$	$\dfrac{FSS/(K-1)}{RSS/(n-K)}$
WITHIN TREATMENTS	$n-K$	RSS	$RSS/(n-K)$	
TOTAL	$n-1$	TSS		

where $n = n_1 + ... + n_K$, and

$$\mathbf{TSS} = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\left(y_{kj} - \overline{y}_{..}\right)^2 \qquad \mathbf{RSS} = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\left(y_{kj} - \overline{y}_{k}\right)^2 \qquad \mathbf{FSS} = \sum_{k=1}^{K} n_k\left(\overline{y}_{k} - \overline{y}_{..}\right)^2 \qquad (3.1.12)$$

where, as above **TSS** is the **total** sum-of-squares (i.e. total deviation from the overall data mean $\overline{y}$) **RSS** is the **residual** sum-of-squares (i.e. sum of deviations from individual sample means $\overline{y}_k$, $k = 1, ..., K$) and **FSS** is the **fitted** sum-of-squares (i.e. weighted sum of deviations of sample means from the overall data mean, with weights equal to number of data points in the individual samples) Note that using (3.1.12), it can be directly verified that $TSS = FSS + RSS$. If the F statistic is calculated in this way, and compared with an F distribution with parameters $K - 1$, $n - K$, the hypothesis that all the individual samples have the same mean can be tested.

EXAMPLE In a bioinformatics context, it is often of interest to compare the structures of different parts of the genome. In this example, three genomic segments were used to studied in order to discover whether the distances (in kB) between successive occurrences of a particular motif were substantially different. Several measurements were taken using for each segment;

	Method		
	SEGMENT A	SEGMENT B	SEGMENT C
	42.7	44.9	41.9
	45.6	48.3	44.2
	43.1	46.2	40.5
	41.6		43.7
			41.0
Mean	43.25	46.47	42.26
Variance	2.86	2.94	2.66

For these data, the ANOVA table is as follows;

Source	D.F.	Sum of squares	Mean square	F
SEGMENTS	2	34.1005	17.0503	6.11
Residual	9	25.1087	2.7899	
Total	11	59.2092		

and the F statistic must be compared with an $F_{2,9}$ distribution. For a significance test at the 0.05 level, F must be compared with the 95th percentile (in a **one-sided** test) of the $F_{2,9}$ distribution. This value is 4.26. Therefore, the F statistic **is** surprising, given the hypothesized model, and therefore there is evidence to reject the hypothesis that the segments are identical.

3.1.6 TWO-WAY ANOVA

One-way ANOVA can be used to test whether the underlying means of several groups of observations are equal Now consider the following data collection situation Suppose there are K treatments, and L groups of observations that are believed to have different responses, that all treatments are administered to all groups, and measurement samples of size n are made for each of the $K \times L$ combinations of treatments $\times$ groups. The experiment can be represented as follows: let y_{klj} be the jth observation in the kth treatment on the lth group, then

$$y_{klj} = \beta_k + \delta_l + \varepsilon_{klj}$$

for $k = 1, ..., K$, $l = 1, ..., L$, and again $\varepsilon_{klj} \sim N\left(0, \sigma^2\right)$. This model assumes that $Y_{kj} \sim N\left(\beta_k + \delta_l, \sigma^2\right)$ and that the expectations for the different samples are different. We can view the data as a 3

dimensional-table comprising K columns and L rows, with n observations for each column $\times$ row combination, corresponding to a sample.

It is possible to test the hypothesis that each **treatment**, and/or that each **group** has the same mean, that is, the two null hypotheses

$$H_0 \quad : \quad \beta_1 = \beta_2 = ... = \beta_K$$
$$H_0 \quad : \quad \delta_1 = \delta_2 = ... = \delta_L$$

against the alternative H_1 :not H_0 in each case. For these tests, a **Two-way Analysis of Variance (ANOVA)** F-test may be carried out. The Two-Way ANOVA table is computed as follows

Source	D.F.	Sum of squares	Mean square	F
TREATMENTS	$K-1$	FSS_1	$FSS_1/(K-1)$	$\dfrac{FSS_1/(K-1)}{RSS/(R+1)}$
GROUPS	$L-1$	FSS_2	$FSS_2/(L-1)$	$\dfrac{FSS_2/(L-1)}{RSS/(R+1)}$
Residual	$R+1$	RSS	$RSS/(R+1)$	
Total	$N-1$	TSS		

where $N = K \times L \times n$, $R = N - L - K$. and again

$$TSS = FSS_1 + FSS_2 + RSS.$$

In the table below, there are $K = 3$ Treatments, and $L = 6$ Groups, and $n = 1$

	I	II	III	GROUP totals
1	0.96	0.94	0.98	2.88
2	0.96	0.98	1.01	2.95
3	0.85	0.87	0.86	2.58
4	0.86	0.84	0.90	2.60
5	0.86	0.87	0.89	2.62
6	0.89	0.93	0.92	2.74
TREATMENT totals	5.38	5.43	5.56	16.37

There are two natural hypotheses to test; first, do the TREATMENTS differ, and second, do the GROUPS differ ?

Two-way analysis of variance: can be used to analyze such data. Given two sources of variation the data can be thought of as a table with the rows and columns representing these two sources . Two-way analysis of variance studies the variability due to the GROUPS effect (here, variability in the specimens) and the variability due to the "column" effect (variability in the TREATMENTS), and calibrates them against the average level of variability in the data overall. Having performed the appropriate calculations, the results are displayed in an ANOVA table. For example, for the data above

Source	D.F.	Sum of squares	Mean square	F
GROUP	5	0.040828	0.0081656	31.54
TREATMENT	2	0.002878	0.001439	5.57
Residual	10	0.002589	0.0002589	
Total	17	0.046295		

The two F statistics can be interpreted as follows; the first ($F = 31.54$) is the test statistic for the test of equal means in the **rows**, that is, that there is no difference between **GROUPS**. This statistic must be compared with an $F_{5,10}$ distribution (the two degrees of freedom being the entries in the degrees of freedom column in the specimens and residual rows of the ANOVA table). The 95th percentile of the $F_{5,10}$ distribution is 3.33, and thus the test statistic is **more extreme** than this critical value, and thus the hypothesis that each specimen has the same mean can be **rejected**.

The second F statistic, ($F = 5.57$), is the test statistic for the test of equal means in the **columns**, that is, that there is no difference between **TREATMENTS**. This statistic must be compared with an $F_{2,10}$ distribution (the two degrees of freedom being the entries in the degrees of freedom column in the methods and residual rows of the ANOVA table). The 95th percentile of the $F_{2,10}$ distribution is 4.10, and thus the test statistic is **more extreme** than this critical value, and thus the hypothesis that each method has the same mean can be **rejected.**